第 35 回

相関係数

講師
湯浅 弘一

まず、ここまでの学習を振り返りましょう。

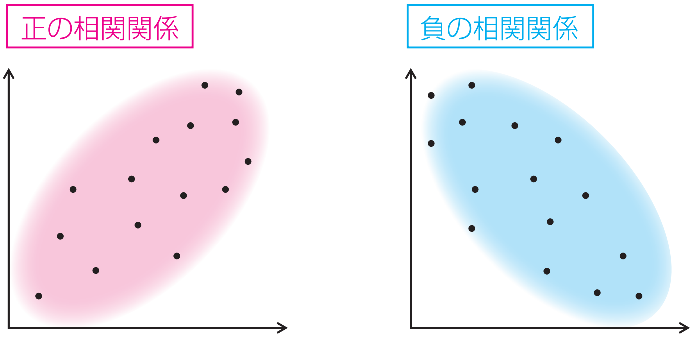
☆ 平均値 ＝（データの値の合計）÷（データの個数）

☆ 偏差 ＝（個々のデータの値）－（平均値）

☆ 分散 ＝（偏差）^{2} の平均値

☆ 標準偏差 ＝ 分散の「正の平方根」（$\sqrt{\text{分散}}$）

☆ 相関関係 … あるデータが変化すると，それに伴ってもうひとつのデータも変化する傾向があること。

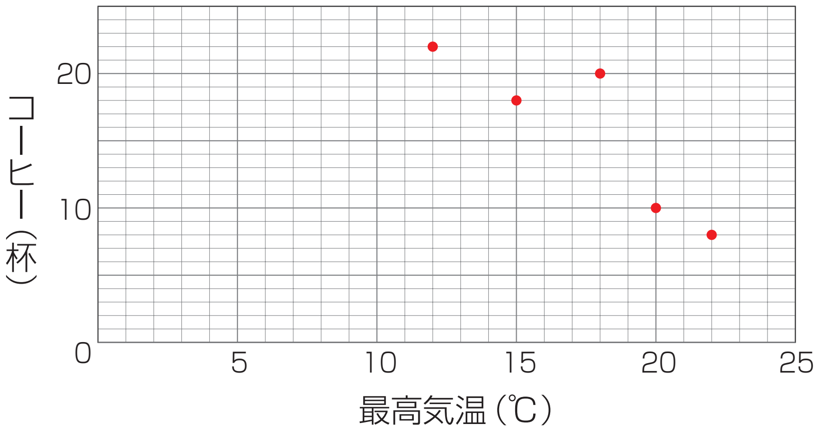


1 共分散

ある５日間の最高気温と，あるカフェのコーヒーの販売数を表にしたものです。

	月	火	水	木	金
最高気温（℃）	20	18	22	15	12
コーヒー（杯）	10	20	8	18	22

これを散布図にすると，次のようになります。



右下がりの楕円になっているように見えますが・・・

相関関係がどのぐらい強いのかを数値で表したものが「共分散」です。

分　散 ・・・１種類のデータの散らばり具合を数値で表す。
共分散 ・・・２種類のデータの相関関係を数値で表す。

共分散は，次のような式で求めます。

{(x の偏差) × (y の偏差) の総和 } ÷ (データの個数)

つまり，偏差の積の平均値です。

共分散が正の値のときは，正の相関関係が，

共分散が負の値のときは，負の相関関係があります。

例 題

最高気温とコーヒーの販売数の共分散を求めなさい。

	月	火	水	木	金
最高気温（℃）	20	18	22	15	12
コーヒー（杯）	10	20	8	18	22

【解 説】

共分散 ＝ {(x の偏差) × (y の偏差) の総和 } ÷ (データの個数)

まず，それぞれの日の偏差を求めます。

偏差は（個々のデータの値）－（平均値）で求められましたね。

	月	火	水	木	金	合 計	平均値
最高気温（℃）	20	18	22	15	12	87	17.4
最高気温の偏差	2.6	0.6	4.6	−2.4	−5.4	0	
コーヒー（杯）	10	20	8	18	22	78	15.6
コーヒーの偏差	−5.6	4.4	−7.6	2.4	6.4	0	

続いて，それぞれの曜日の偏差の積を求めます。

	月	火	水	木	金	合 計	平均値
最高気温（℃）	20	18	22	15	12	87	17.4
最高気温の偏差	2.6	0.6	4.6	−2.4	−5.4	0	
コーヒー（杯）	10	20	8	18	22	78	15.6
コーヒーの偏差	−5.6	4.4	−7.6	2.4	6.4	0	
偏差の積	−14.56	2.64	−34.96	−5.76	−34.56		

５日間の平均を求めるので，

$(-14.56 + 2.64 - 34.96 - 5.76 - 34.56) \div 5 = -87.2 \div 5 = -17.44$

共分散は－ 17.44 と求まりました。負の値なので，負の相関関係がありそうですね。

2 相関係数

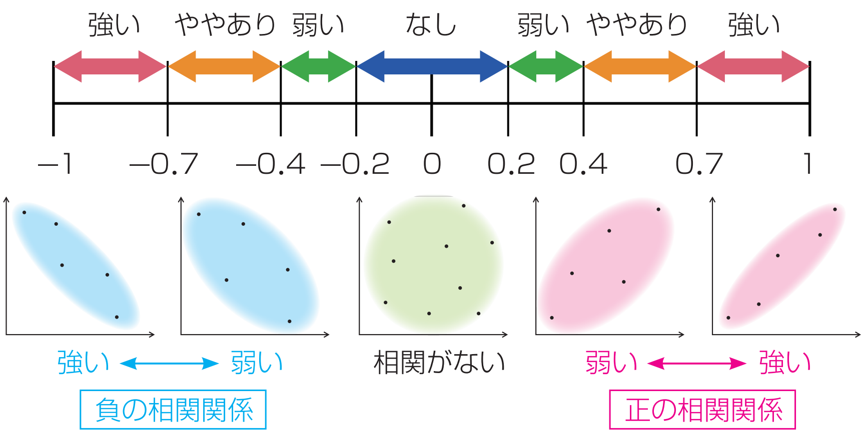
相関関係の強弱を数値で表したものが「相関係数」です。

相関係数はこのような式で求めます。

$$\text{相関係数 } r = \frac{x \text{ と } y \text{ の共分散}}{(x \text{ の標準偏差}) \times (y \text{ の標準偏差})}$$

相関係数 r は，必ず $-1 \leqq r \leqq 1$ となります。

相関係数は以下のように，相関関係の強さを表しています。



例 題

あるカフェの５日間の，ハンバーグとパフェの売り上げについて，相関係数を求めなさい。
ただし，小数第３位を四捨五入して小数第２位まで求めること。

	月	火	水	木	金
ハンバーグ（個）	3	4	4	5	5
パフェ（個）	2	4	5	4	5

【解 説】

$$\text{相関係数 } r = \frac{x \text{ と } y \text{ の共分散}}{(x \text{ の標準偏差}) \times (y \text{ の標準偏差})}$$

まず，この公式の分母の部分，それぞれの標準偏差を求めましょう。

①標準偏差 ＝ $\sqrt{\text{分散}}$

②分散 ＝（偏差）^{2} の平均値

③偏差 ＝（個々のデータの値）−（平均値）

③②①の順に求めていきます。

	月	火	水	木	金	合計	平均値	標準偏差
ハンバーグ（個）	3	4	4	5	5	21	4.2	
ハンバーグの偏差	−1.2	−0.2	−0.2	0.8	0.8	0		
(ハンバーグの偏差)^2	1.44	0.04	0.04	0.64	0.64	2.8	0.56	0.75
パフェ（個）	2	4	5	4	5	20	4	
パフェの偏差	−2	0	1	0	1	0		
(パフェの偏差)^2	4	0	1	0	1	6	1.2	1.10

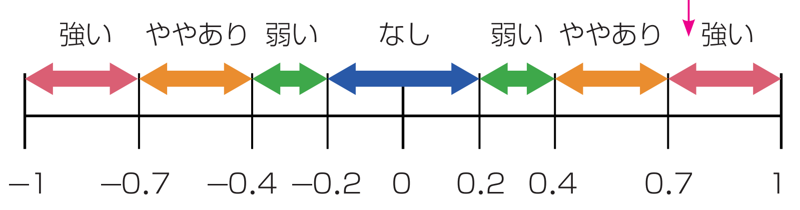
ハンバーグの売り上げの標準偏差は 0.75
パフェの売り上げの標準偏差は 1.10
これで，

$$\text{相関係数}\ r = \frac{x\ \text{と}\ y\ \text{の共分散}}{(x\ \text{の標準偏差}) \times (y\ \text{の標準偏差})}$$

この公式の分母の部分は代入できます。
次に，分子の部分を求めます。
共分散は，偏差の積の平均値でしたね。

	月	火	水	木	金	合計	平均値	標準偏差
ハンバーグ（個）	3	4	4	5	5	21	4.2	
ハンバーグの偏差	−1.2	−0.2	−0.2	0.8	0.8	0		
(ハンバーグの偏差)^2		0.04	0.04	0.64	0.64	2.8	0.56	0.75
パフェ（個）	2	4	5	4	5	20	4	
パフェの偏差	−2	0	1	0	1	0		
(パフェの偏差)^2	4	0	1	0	1	6	1.2	1.10
偏差の積	2.4	0	−0.2	0	0.8	3	0.6	

ですから，公式の分子の部分は，
$(x$ と y の共分散$) = (2.4 + 0 - 0.2 + 0 + 0.8) \div 5 = 3 \div 5 = 0.6$
となります。
よって，
相関係数 $r = (0.6) \div (0.75 \times 1.10) = 0.6 \div 0.825 \fallingdotseq 0.73$
相関係数は約０．７３となるので，強い相関関係があるといえます。



やってみよう！

５人の生徒の国語と社会の点数の、相関係数を求めなさい。

	Aさん	Bさん	Cさん	Dさん	Eさん
国　語（x点）	3	4	4	5	5
社　会（y点）	2	4	5	4	5

【答 え】

まずは、共分散を求めるために偏差を求めます。

	Aさん	Bさん	Cさん	Dさん	Eさん	合 計	平均値
国　語（x点）	3	4	4	5	5	21	4.2
国語の偏差	−1.2	−0.2	−0.2	0.8	0.8	0	
社　会（y点）	2	4	5	4	5	20	4
社会の偏差	−2	0	1	0	1	0	

国語の平均点は 4.2 点。社会の平均点は４点ですから

共分散は　$\{(-1.2)\times(-2)+(-0.2)\times 0+(-0.2)\times 1+0.8\times 0+0.8\times 1\}\div 5=0.6$

次に，国語の標準偏差と社会の標準偏差を求めます。

分散は，偏差２乗の平均でしたね。

ですから，国語の分散は，

$\{(-1.2)^2+(-0.2)^2+(-0.2)^2+0.8^2+0.8^2\}\div 5=0.56$

標準偏差は，その正の平方根なので，$\sqrt{0.56}$

同様に，社会の分散は，

$\{(-2)^2+0^2+1^2+0^2+1^2\}\div 5=1.2$

標準偏差は，その正の平方根なので，$\sqrt{1.2}$

これらの数値を　相関係数 $r=\dfrac{x\text{ と }y\text{ の共分散}}{(x\text{ の標準偏差})\times(y\text{ の標準偏差})}$ に代入して，

相関係数 $r=\dfrac{0.6}{\sqrt{0.56}\times\sqrt{1.2}}=\dfrac{0.6}{\sqrt{0.56\times 1.2}}=\dfrac{0.6}{\sqrt{0.672}}=0.73\cdots$

と求まります。

目安の 0.7 を超えているので，強い相関が見られます。

つまり，国語の得点が高いと，社会の得点も高い傾向が強いということです。